Explainable AI for Recommender Systems

Jennifer Swaminathan
University of Neuchatel,
Neuchatel, Switzerland
jennifer.swaminathan@unine.ch

Supervisors: Mira El Kamali & Quentin Meteier
University of Fribourg
Fribourg, Switzerland
mira.elkamali@hefr.ch quentin.meteier@hefr.ch

***Abstract* –This paper reviews previous Recommender Systems using Explainable Artificial Intelligence (AI). We present the existing systems, models, and modalities used and their effect on the acceptance rate. From the comparison, it can be seen some criteria need to be followed to obtain these Explainable AI for Recommender Systems. In the end, we define a taxonomy of explainable AI for recommender systems.**

I. I.INTRODUCTION

Recommender Systems provide solutions to information overload, e.g., in an eCommerce website, presenting suggestions to the user. Recommender Systems are used to find relevant products to the user, by finding similar products that he has purchased (Collaborative filtering) or by finding products that match the user's profile (content-based) [1]. To increase the accuracy of the prediction, deep learning technologies are being used. These have resulted in black boxes which do not have interpretability. To alleviate these problems, we use Explainable AI to obtain the reason for these predictions. Explainability can be achieved using a transparent design or post-hoc methods [1]. Thus, we get Explainable AI for Recommender Systems which provides clear information to the user as to why an item has been predicted to increase the trust. To make these Recommender Systems explainable, there are different modalities used to explain the reason for the output; for example, in [2], textual explanations are given, and [3] uses visual description by highlighting the point of interest in a fashion image, hybrid explanations provide a textual explanation along with an image [4], and social explanations [5], which includes social factors, such as friends interest on an item. This literature review also explores Recommender Systems in Energy management [6, 7, 8, 9] models. The recommendations presented to the user also include a persuasive explanation to bring about behavior changes in the users by providing them with ecological or economic facts to help reduce energy consumption.

In this paper, we want to answer the following questions:

Q1: Which modalities can convey explanations to the end-user in Explainable Recommender Systems?

Q2: How (persuasive) explanations can increase acceptance of Explainable Recommender Systems?

The structure of this literature review is as follows. Section 2 presents the definitions of some terms such as Recommender Systems, Explainable AI, and Explainable Recommender Systems. Section 3 presents the methods and results used to answer Q1 and Q2. These results include a comparison of the criteria used in different models, and a synthesis table with discussion. In section 4, a design of a taxonomy for Explainable Recommender Systems is explained. Section 5 contains the conclusion and limitations of this paper.

II. DEFINITIONS

Artificial Intelligence has gained importance in various fields. Humans have started trusting these AIs in autonomous vehicles, aircrafts using autopilot, medical domains to find a mismatch in genes, finance domain to find defaulters in banks, and various other fields. These AIs use deep learning to obtain their results. These deep learning technologies use multiple levels of neural network layers to get the results, which are not interpretable by the developers. This leads to dependency on AI to obtain our goal. For increased accuracy in the model, the explainability of the model decreases, which leads to a tradeoff between accuracy and explainability.

With sensitive data being handled, how safe is it to rely on the AI, which is a black box? If there is transparency in how the calculations are computed, humans have more control over the system. For this, we need Explainable AI(XAI). These XAI make the systems interpretable by making the black box transparent or using post-hoc methods, where the black box and an Explainable sub-system are used to explain. These models must lead to the interpretability of the AI systems, thus increasing the end user's trust.

In the current age of information overload, users are overwhelmed by the available content. On an eCommerce site, with millions of products available, how do we make users buy more to increase the revenue? This is possible by providing recommendations that can persuade them to buy more. These recommendations are obtained by Recommender Systems, which builds user profiles from their activity on the site. Each user has a unique profile, so not all users are provided with the same recommendation.

There are two main approaches to building Recommender Systems, content-based and collaborative filtering. Content-based Recommender systems are systems where a user profile is created by obtaining information through the user's interaction on the website to make predictions related to what the user likes. Collaborative filtering [3] uses users' historical rating of items and similarity measures to find the neighborhood of similar users and items to predict. Other types of recommender systems are Context-aware systems which consider the context before prediction, e.g., geo-location of the user, the mood of the user, etc. These methods are used in [2, 4, 7]. Hybrid recommender systems use combinations of the algorithms as in [5, 10]. But to increase the accuracy of these recommendations, deep learning technologies are being used, leading to uninterpretable results.

When the recommendation presented to the end-user contains an explanation, the user's trust increases, and improves the acceptance rate. These are known as Explainable Recommender Systems (ERS).

Why is ERS essential? [11] These systems provide transparency to the prediction, which increases users' trust in these systems and leads to more revenue for the developers. It gives satisfaction to the users. These explanations can be persuasive, and it increases the acceptance of the recommendation by increasing the speed of choosing an item instead of searching for the item.

Why do we need ERS? To justify the prediction, discover patterns, control which items can be presented to the user, and identify errors and correct them.

III. EXPLAINABLE RECOMMENDER SYSTEMS

A. *Methods*

We started our literature review on Explainable AI in recommender systems by searching Scopus with (TITLE-ABS-KEY (explainable AND recommended AND systems) AND TITLE-ABS-KEY (energy AND efficiency)), which resulted in only five papers. So, we decided to remove energy efficiency and selected 12 interesting papers to include broader applications to work on our literature review; along with the papers with energy efficiency, a reduction in energy consumption is an essential topic in recent times. After selecting the papers, we tried to synthesize them into standard criteria to be able to compare them with one another. While doing so, we were able to find six criteria that were present in the models. Namely, the W problem, which explains what the Explainable Recommender System model can solve, Recommender Models, which are used to predict the item which we want to present to the user; explanation modalities, the style the user receives the explanation for the predicted item, Input data these are additional data required to produce explanation to the user, Computing platforms whether the computation of the predictions are done on cloud or on edge servers and Evaluation techniques the different metrics used to evaluate the model to show that the explanation helps in obtaining better accuracy of prediction.

B. *Results*

In this subdivision, we present the results of our literature review, which are the different criteria followed by the Explainable Recommender Systems models.

1) *W-Problem*

Explainable Recommender systems are used to answer W problems [1]. These Ws are what, where, who, when, whether, and why. WHAT answers which item is being recommended to the user. WHERE answers location details to be considered to present the item, e.g., the user's current location is deemed to recommend a restaurant near him, WHO answers the social aspect for the recommendation as in [5, 10] they consider the number times a song is listened to in the community and present to the user, WHEN considers time, in [4] they consider dynamic changes in the user to include in predicting the items. WHETHER considers if the user likes the recommendation or not, in [2], the author uses this criterion so that the model can update the user's profile. All Explainable Recommender Systems try to answer WHY, which explains why the model selected this item to present to the user.

2) *Recommendation models*

In ERS models, different components are embedded in the model to obtain explainability. The authors in the papers try to solve specific questions. In [4], the authors consider the dynamic nature of users, that is, how users' interest in a topic change over a period. If the model keeps recommending an item that the user finds boring or not interested, his trust in the model decreases. To take this into account, the authors consider using Time-aware GRU (gated recurrent unit), which provides weights to the concepts (fundamental interest) based on time. They use CNN (convolution Neural Networks) on the user reviews to obtain the key concept the user mentions in his reviews. The weight on these topics decreases over time so that the concept that the user has preferred in the past week has more weight than the item he chose two years ago. This methodology results in item prediction and an explanation that highlights the item review with the key concept as an explanation to the user.

In [2], the author tries to integrate the user's current mood to obtain items to recommend to the users. The model creates a user profile based on the previous activities, a global profile. The agent recommends an item from the user profile, for which the user either accepts or gives feedback. The Microsoft concept graphs feedback obtains the concept to update the local user's profile. For example, if the user is not in the mood for the comedy genre, he can provide feedback to the model, as shown in Figure 1, which updates the local profile by removing that genre(concept) from the model, hence only updating the local profile and not the global profile. Comedies are not entirely removed from the user's profile. The model explains the item by providing the most relevant item review based on the key concept of the user.

In [3], the authors want to include the fact that the users are interested in different features of a fashion image. For example, in a fashion image, some users might be interested in the collar shape, others in the length of the sleeve or if a pocket is present. To take this into account, the model uses fine-grained visual preference modeling where the image of the fashion is integrated with user reviews to obtain the attention areas to form a multimodal attention network, which tries to pinpoint the location of interest in the image.

[5, 10] uses ERS in the field of music recommendation. In [5], the authors want to explain why a song is recommended to

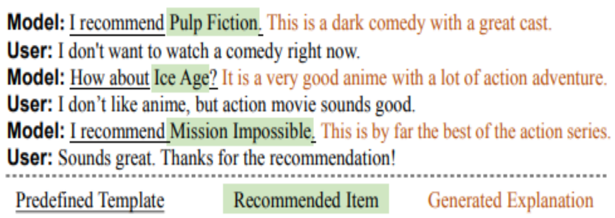

Figure 1: Explainable Conversation Recommendation

the user, whereas in [10], the authors want to provide a recommendation of songs that have already been listened to by the user. They use the HyPER framework, which uses probabilistic soft logic, and the Hybrid Recommender system. HyPER is provided with a set of rules, and it calculates the probability between these rules to choose the most probable item as an explanation to present to the user, as shown in Figure 2.

$$\textsc{SimUsers}_{CF}(u_1, u_2) \wedge \textsc{Listens}(u_1, a) \Rightarrow \textsc{Listens}(u_2, a)$$
$$\textsc{SimArtists}_{CF}(a_1, a_2) \wedge \textsc{Listens}(u, a_1) \Rightarrow \textsc{Listens}(u, a_2)$$
$$\textsc{SimArtists}_{last.fm}(a_1, a_2) \wedge \textsc{Listens}(u, a_1) \Rightarrow \textsc{Listens}(u, a_2)$$
$$\textsc{SimArtists}_{content}(a_1, a_2) \wedge \textsc{Listens}(u, a_1) \Rightarrow \textsc{Listens}(u, a_2)$$
$$\textsc{HasTag}(a_1, t) \wedge \textsc{HasTag}(a_2, t) \wedge \textsc{Listens}(u, a_1) \Rightarrow \textsc{Listens}(u, a_2)$$
$$\textsc{SimFriends}(u_1, u_2) \wedge \textsc{Listens}(u_1, a) \Rightarrow \textsc{Listens}(u_2, a)$$
$$\textsc{PopularArtist}(a) \Rightarrow \textsc{Listens}(u, a)$$
$$\neg\textsc{Listens}(u, a)$$

Figure 2:Rules used in HyPER model

In [7, 8, 9], use the EM3(energy management) model, which consists of a knowledge base that contains all the user's activities when he uses the appliances and at which ambient conditions he turns on/off devices. As shown in Figure 3, the model takes real-time measurements from the sensors. This model uses a context-aware recommender system as it looks at the ambient conditions. These are provided to the ATM (Action triggering model). The ATM preprocesses information to the recommendation engine, which decides when to provide a recommendation to the end-user. Here, the user is given a text message, and the actuators turn off the appliance based on the user's feedback. [7] uses Fusion-based recommender systems, where the ATM compares the user's preference of the environment and the current sensor readings to provide an SMS on the Telegram App on the user's mobile, with the recommendation and an explanation. The users can accept, reject, or ignore this recommendation. This feedback is used to update the user's profile. If the recommendation is accepted, similar explanations with reasoning are provided to the user. If

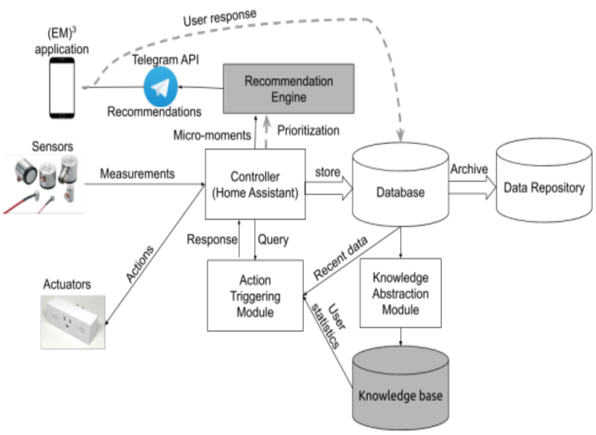

Figure 3:EM3 architecture

the recommendation is ignored or rejected, then other means of explanation are used.

Knowledge graphs [12, 13] use the relationship between users and items to build graphs to form a rich knowledge to explain to the user. The relationship between the entities can be if the user purchased the item or viewed the item which can be used as an explanation to the user. In [13], the author integrates the user's sentiments in the review to aid in explanation and help in obtaining more accuracy in the prediction. These sentiments can include if the user is satisfied or dissatisfied with the product. In Figure 4 we can see how this model can consider the user's sentiment to help obtain an accurate prediction. For example, Jane and Leon are satisfied with the phone case, so Jane can be recommended a product that Leon is satisfied with.

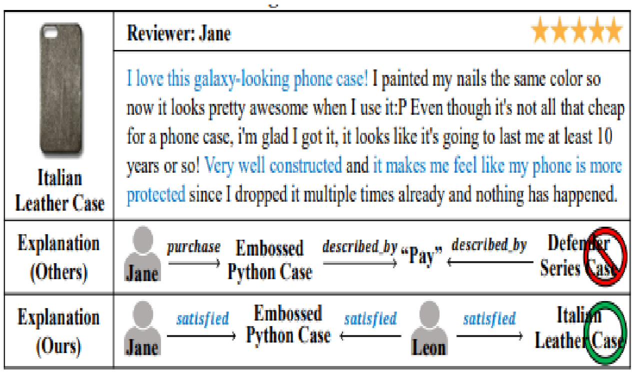

Figure 4: Knowledge Graph explanation with sentiment analysis

3) Explanation modalities

An explanation of a product recommendation is key to increasing the trust in the system. So, the model needs to make it interpretable. In this section, we answer Q1.

Different explanation styles that can be provided to the users are: i) Textual explanation, ii) Visual explanation, iii) Item-based or user-based explanation, iv) Social-based explanation, and v) Hybrid explanations.

Textual explanations are used in [2, 5, 7]. In [2], the model produces textual explanations from predefined templates adding the recommended item as shown in Figure 1. The generated explanation. The model integrates feedback from the user and updates the user profile to present the user with accurate items.

In [7, 8, 9], the textual explanation presented to the users is persuasive, including facts about ecological and economic impact, leading to behavior changes. The recommendation provided to the user includes facts about user presence in a room or the general context; for example, if the outside temperature is pleasant, the recommendation can be to open the window and turn off the AC. This is integrated with a persuasive explanation. This can be based on ecological or economic reasoning. An example of ecological reasoning is explaining carbon footprint. At the same time, an explanation of economic reasoning provides cost savings by turning off the AC

Visual explanations are used in [3, 4]. In [3], the model highlights a personalized point of interest in the image as an explanation to the user. It considers users' reviews as side information to obtain the concept in the image and attention mechanism on the image to locate the idea and then explains as shown in Figure 5. From the user's reviews for previous items he has purchased, the model finds the concept that the user is interested in and pinpoints that part on the image. For example, for the target item of the shoe, the user has previously written about toe fit, and when the image is recommended to the user, the model highlights the toe to the user in VECF.

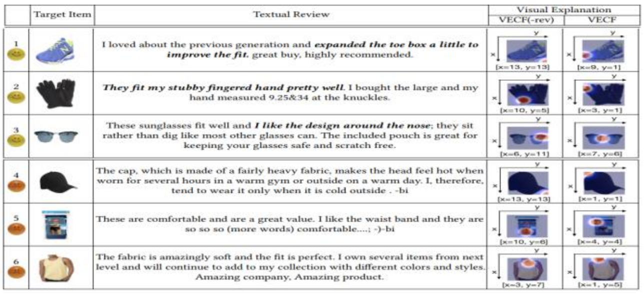

Figure 5: Visual Explanation

Hybrid explanation [4] uses both textual and visual to provide a clear explanation. This model uses Convolutional Neural Networks to obtain concepts in the users review and find the users' dynamic interests. Time-aware GRU is used to form user profiles. The explanation highlights the more appropriate item review to the user based on the concept he is most interested in. In Figure 6, we can see that the first user gets Review No2, as he talks about 'quality' in his previous reviews, and that concept is highlighted as the explanation of the item, and for the second user, Review 4 and 6 are highlighted along with the key concept 'fits'.

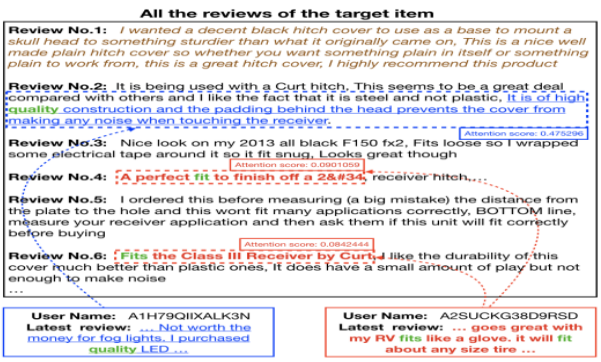

Figure 6:Hybrid explanation in DER

In [13], an interactive UI is built to present the recommendation to the user. It uses textual explanation and feature extraction to explain to the user. Here the user's review is used to obtain the feature that the user talks about and explain the reason for the prediction, as shown in Figure 7.

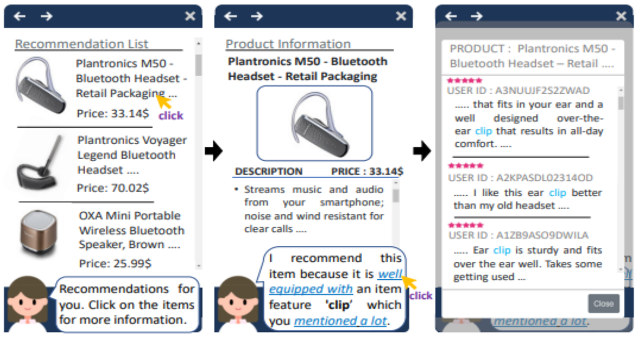

Figure 7:Explanation through UI

Social explanation and item/user explanation are used in [5, 10]. Here, the model provides textual explanations and social item factors to provide persuasive explanations to the users, as shown in Figure 8.

Factor	Style	We recommend *Shake It Off* by *Taylor Swift* because:
Personal factor	P-first	Exactly *five years* have passed since you listened to it for the first time.
	P-last	Exactly *three years* have passed since you listened to it last.
	P-together	Around the same time, you frequently listened to it and *Applause* by *Lady Gaga*, which you listened to just now.
	P-total	You will have played it *100* times when you listen to it next.
Social factor	S-total	Its total play count by all users reached *one million*.
	S-unique	The number of unique users who listened to it reached *100 thousand*.
	S-favorite	The number of users who added it to their Favorites reached *10 thousand*.
Item factor	I-release	Exactly *five years* have passed since it was released.
	I-live	*Today, Taylor Swift* performed it at a live concert.

Figure 8:User/Item/Social Factor explanation

4) *Input data*

For explainability, in the recommender system, we embed additional data to provide the interpretability of the recommendations. Different types of information can be used, like the user's profile, review and rating, and sensor readings. [2, 3, 4] uses user reviews, and item reviews to provide accurate predictions and explanations. User profiles are used in [7] to give a personalized explanation to the user to obtain satisfaction. Sensor reading and ambient conditions [7, 8, 9] are used as side information to explain to the user.

But, when there is a lot of side information considered, there needs more memory, which gets computationally expensive and increases the latency to obtain predictions and explanations to present to the user. Adding more variables to the model reduces the model's accuracy, as found in [12]. Their evaluation reduced computation time by 80% by compressing the side information.

5) *Computing platforms*

Since a large amount of computation needs to be done to compute the prediction score for the users, developers try to use cloud platforms for analysis. But these platforms are not very trustworthy, as users' preferences and profiles are sent over the cloud for computation. This leads to privacy issues. For example, if the user's presence is being hacked in the energy model, it could lead to information leakage. Also, other problems faced by using cloud platforms are that models are dependent on the internet to connect to the server for computation. When there is a failure in the cloud servers, information is lost. There is also a delay in time to compute the prediction. To deal with this problem in [8], they use edge servers to compute the sensitive data and avoid data leakage. Therefore, deciding which platform to perform calculations can be application-dependent for ERS models. For example, to recommend music with a lot of computation, cloud platforms could be helpful. Whereas in the Energy model, the user's location is constantly monitored, hence analysis on edge servers could be a solution. If cloud platforms are being used, proper encryption needs to be performed on sensitive data before the data is transferred.

6) Evaluation Techniques

In this section, we answer Q2. To evaluate the models, the authors perform various evaluation techniques. They can be offline, online, or a user study [1].

In **offline mode**, the historical data of the user's interaction is already present, e.g., the items that the user has purchased. Offline evaluations are also known as retrospective evaluations or dead data. Generally, 80% of the data is used to train the model and 20 % to test the model. The model should be able to predict the items that the user had purchased.

Online evaluations are known as prospective evaluations or live experiments. In an online test case, many users and interactions are required in real-time to check if the provided recommendation is accepted by the users.

In **User studies**, users are made to interact with the model. With these interactions, users are provided with recommendations and explanations, and they are asked to evaluate the model. The problem with user studies is that not many participants are present. For example, in [10], a user study is performed to evaluate the recommendation of repeat consumption explanation. Still, the participants were only considered from Japan, leading to biased results. User studies are preferred to online as users can provide direct feedback.

In the models studied qualitative and quantitative analyses are performed. Quantitative analyses calculates the Hit ratio, NDCG (Normalized Discounted Cumulative Gain), MRR (Mean Reciprocal Rank), Precision, and Recall to find if the predicted items match the ground truth. In qualitative analysis, the explanation provided by the model and baseline models are presented to the user studies, and the quality of explanation is asked to be rated by the participants of the user studies. Some ERS models perform ablation analysis as there are not many models to compare their features.

Now we discuss the evaluations techniques used in the ERS models. In [2], offline evaluation methods are used to evaluate the explanations. They used 80% of their dataset for training, 10% for hyperparameters setting, and 10% for tests. They used the first sentence of the review as the ground-truth explanation, which they wanted to predict and use as an explanation. They used HR(hit ration) (for recommendation accuracy), NDCG(normalized discounted cumulative gain)(is used to evaluate the ranking position of the item), and MRR(mean reciprocal rank) to assess the accuracy of their prediction model. They use ablation analysis to study how the view with global and local components and how the user's concept inclusion in the feedback can produce better predictions. They can see that a model with only a Local profile performs better than their model since it considers only the current concepts of the user. [3, 13] also uses an offline dataset with a 70-30 ratio to train and test their model. Quantitative evaluation (Hit Ratio and Normalized Discounted Cumulative Gain) showed promising results. [2, 3, 13] use ablation analysis since no baseline model used these explanation techniques. They could see that their model could always produce better predictions and provide a better explanation.

In DER [4], the authors compare a baseline model (NARRE) to perform qualitative and quantitative evaluations, with DER performing better than NARRE in its explanation. NARRE always highlights the same review for all users for an item, compared to DER, which provides a personalized basis compatible with each user.

Energy Model [7] used user studies on office-based scenarios; the participants were provided with office-based situations and recommendations. The model was updated and provided feedback to the participants. Three recommendation scenarios were presented to the user, as shown in Figure 9. In Scenario I, the user was provided with the recommendation without any explanation. In Scenario II, the user was provided with a persuasive explanation, providing reasons for economic and ecological benefits. In Scenario III, the user was provided with a persuasive explanation and reasoning that the user was not present in the room or that the temperature outside was pleasant to open windows. The acceptance rate increases by adding persuasive reasoning based on user profile with ecological or economic reasoning. Using user studies, the authors found that using ecological and economic reasoning, there was an increase of 19% in the acceptance ratio.

In [10], a user study was performed to find which style of explanation was preferred. A total of 9 explanations were presented to the user, as shown in Figure 8 and the user was asked to provide their preferred one. P- total, that is, an explanation (*you will have played SONG A 100 times when you listen to it next.*) was preferred to other styles of explanation.

In [5], they perform a user study on the online forum 'last.fm'. They evaluate the number of explanations that a user would require to persuade. They found a mean of 3 to 4 was

Figure 9:Scenarios of explanation in Energy Model

TABLE 1: Synthesis of Explainable Recommender System

	Towards Conversational Explainability	Personalized Fashion Recommendation with Visual Explanations based on Multimodal Attention Network	Dynamic Explainable Recommendation Based on Neural Attentive Models	Delivering explainable and personalized recommendations for energy efficiency	Smart fusion of sensor data and human feedback for personalized energy-saving recommendations	Intelligent edge-based recommender system for internet of energy applications	Explainable Recommendation for Repeat Consumption	Personalized Explanations for Hybrid Recommender Systems	Reinforcement Learning over Sentiment-Augmented Knowledge Graphs towards Accurate and Explainable Recommendation
Year	2020	2019	2019	2020	2021	2021	2020	2019	2022
Aim	Explanation through multi-turn feedback	Explanation by highlighting point of interest on an image	Explanation considering users current state of mind.	Persuasive explanation for energy management	Persuasive explanation using fusion based RS	Persuasive explanation by performing computation in edge servers to avoid information leakage	Explaning repeat consumption songs based on social, item, personal factors	Provides real time recommendation with 7 textual explanations	Uses sentiment analysis to provide explanation
Concept	Context aware	Collaborative Filtering	Context aware	Context aware	Data fusion	Edge/fog based hybrid RS	Hybrid	Hybrid	Knowledge graph RS
W problems	What, whether,why	why, what	when, what, why	Why , what	Why , what	Why , what	who , what , why	who, what, why	why , what,who
Recommendation models	Multi-task learning framework, multiview feedback integration	Fine grained visual preference modeling , GRU on user review to form multimodal attention network	Time aware GRU-user profile, CNN- user review	Action Triggering model,Knowledge Base, Recommendation engine	stacked LSTM Neural network, for fusion of data	stacked LSTM Neural network	Probabilistic soft Logic, using rule based explanation	Probabilistic soft Logic, using rule based explanation	Uses KG to explain the reason, by using sentiment analysis..
Input data	User / item reviews	User / item reviews	User / item reviews	user profile, sensors	user profile, sensors	user profile, sensors	item/user/social info	item/user/social info	User / item reviews
Computing platform	-	-	-	-	Cloud Servers	Edge Server	-	-	-
Explanation modality	Textual explanation	Visual explanation	Visual explanation	Contextual and textual explanation	Contextual and textual explanation	Home assistant UI, textual /visual explanation	Textual explanation with social and Item, user explanation	Textual explanation with social and Item, user explanation	Interactive UI with Textual explanation
Evaluation Technique	Offline, User study (amazon)	Offline, user study (amazon)	Offline/user study (amazon &yelp)	user studies	Online (EM3)	Online(EM3)	user studies	user studies along with online(lastfm)	Offline datasets (amazon), user studies
Evaluation: Quantitative	HR, NDCG, MRR- for prediction.	HR and NDCG	RMSE- for prediction explanation	Accuracy of prediction	Accuarcy of prediction	Decrease in latency			NDCG, HR,F1, Precision, Recall
Qualitative	user study with experts	Acceptance rate increases with explanation	Personalized explanations increases acceptance rate	Persuasive explanation increases acceptance rate	Persuasive explanation increases acceptance rate	No data loss	P total persuades users	No of explanations and styles are compared	User study- compare explanations between models
Ablation	Yes	Yes	No	No	No	No	No	No	Yes
Human-in-the-loop	Yes	-	-	Yes	Yes	Yes	-	-	Yes

preferred. They evaluated the preference in explanation technique, whether a Venn diagram, dendrogram, or textual explanation were selected. The evaluations showed that the users preferred textual explanations to other explanations.

In [9], the accuracy of prediction is the main criterion evaluated. Online evaluations were performed with a real-time feed of the values of the sensors. There is a delay in obtaining the signals, and hence there is a delay in a recommendation. The prediction accuracy reached a maximum of 90% for all the three scenarios they studied.

In [8], online evaluations were performed on the University campus. They perform computation on the edge server, thus reducing latency, decreasing dependence on the internet, and avoiding privacy issues.

C. Discussion

From the results of our literature review, given in Section III B, we present a synthesis of the models studied in TABLE 1. This synthesis compares the aim of the different models', the recommendation algorithm, if the model has humans in the loop integrated, the W problem answered by each model, additional input data used for explanation, evaluation techniques, and explanation modalities.

To answer our literature question Q1, explanation modalities in each model (highlighted in yellow) are compared. Almost all models use textual explanation. These textual explanations can include user, item, or social factors to make them more persuasive. Others have contextual explanation integrated to form a hybrid explanation. In [3], they use visual explanation to highlight the point of interest to the user.

From the evaluations of the models, we can answer our Q2, i.e., explanations can provide a better acceptance rate through quantitative, qualitative analysis and ablation analysis. From the table, we can see that not all models discuss computing platforms, as they are features that are considered when the model has been deployed. Certain criteria are needed to be satisfied to build an ERS, so we build a taxonomy for ERS in the following section.

IV. DESIGN OF A TAXONOMY OF ERS

We present a taxonomy for Explainable Recommender Systems that developers of ERS model can use to follow while building their model. The criteria are **W-problem**, what kind of problem ERS model will solve, **Explanation Modalities**, which is the style of explanation that will be used to present to the user, **Recommendation Model**, different algorithms used to predict the item to present to the user. **Input Data**, additional data needed for explaining to the user, **Computing platform** where the computation of the algorithm can take place, and finally, the **Evaluation techniques**, which are required to

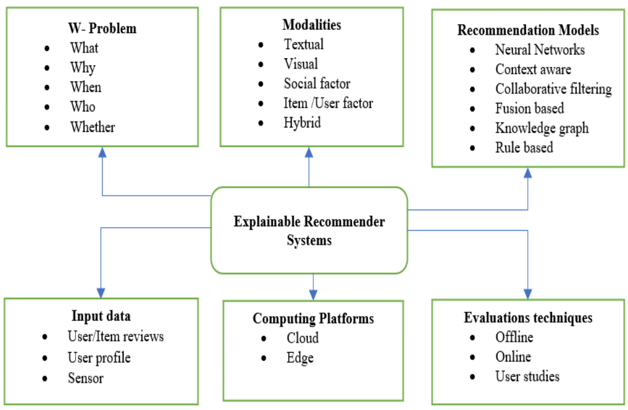

Figure 10:Taxonomy of Explainable Recommender Systems

evaluate the model to obtain accuracy along with interpretability of the model. This taxonomy is presented in Figure 10. There are other taxonomies for Explainable Recommender Systems reviewed in other reviews. In [14], a taxonomy of Explanation styles used is presented. This taxonomy contains only the modalities compared to other criteria included in our taxonomy. In [15], they have created a taxonomy on ERS with the criteria: Motivation, Knowledge, Generation, and Presentation. In [6], a taxonomy for Energy saving Recommender Systems was presented where they consider objectives, main stages, evaluation metrics, Incentive measures, and other criteria.

V. Conclusion

This literature review presents the existing Explainable AI in Recommender Systems. It compares the models and designs a taxonomy that can be followed to build an Explainable Recommender System. Using Explainable AI, we can provide transparency to our model and offer persuasive explanations leading to an increase in acceptance rate. The persuasive explanation can also offer behavioral changes in humans to accept the recommendations and can help with reducing energy consumption, as seen in Energy Model.

This literature review is not extensive as not all the state-of-the-art models are studied. Although our paper does not consider all the criteria present in all the Explainable Recommender Systems models, with this paper, implementers of ERS can have a general overview on what are the essential criteria required to build an ERS and follow the taxonomy to develop their system. In future work, the taxonomy can be enhanced with further criteria included as studied in [15, 14] to build a more comprehensive model.

References

[1] Y. Zhang and X. Chen, "Explainable recommendation: A survey and new perspectives," *Foundations and Trends® in Information Retrieval 14.1,* vol. 1, no. 1, pp. 1-101, 2020.

[2] Z. Chen, X. Wang, X. Xie, M. Parsana, A. Soni, X. Ao and E. Chen, "Towards Explainable Conversational Recommendation," *IJCAI,* pp. 2994-3000, 2020.

[3] X. Chen, H. Chen, H. Xu, Y. Zhang, Y. Cao, Z. Qin and H. Zha , "Personalized fashion recommendation with visual explanations based on multimodal attention network: Towards visually explainable recommendation," *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval,* pp. 765-774, 2019.

[4] X. Chen, Y. Zhang and Z. Qin, "Dynamic Explainable Recommendation Based on Neural Attentive Models," *Proceedings of the AAAI Conference on Artificial Intelligence.,* vol. 33, no. 1, pp. 53-60, 2019.

[5] P. Kouki, J. Schaffer, J. Pujara, J. O'Donovan and L. Getoor, "Personalized Explanations for Hybrid Recommender Systems," *Proceedings of the 24th International Conference on Intelligent User Interfaces,* pp. 379-390, 2019.

[6] Y. Himeur , A. Alsalemi, A. Al-Kababji, F. Bensaali, A. Amira, C. Sardianos, G. Dimitrakopoulos and I. Varlamis , "A survey of recommender systems for energy efficiency in buildings: Principles, challenges and prospects," *Information Fusion,* vol. 72, pp. 1-21, 2021.

[7] C. Sardianos, I. Varlamis , C. Chronis, G. Dimitrakopoulos, A. Alsalemi, Y. Himeur, F. Bensaali and A. Amira , "The emergence of explainability of intelligent systems: Delivering explainable and personalized recommendations for energy efficiency," *International Journal of Intelligent Systems.,* vol. 36, no. 2, pp. 656-680., 2020.

[8] A. Sayed , Y. Himeur, A. Alsalemi , F. Bensaali and A. Amira , ""Intelligent edge-based recommender system for internet of energy applications," *IEEE Systems Journal,* 2021.

[9] I. Varlamis, C. Sardianos , C. Chronis , G. Dimitrakopoulos, Y. Himeur , A. Alsalemi , F. Bensaali and A. Amira , "Smart fusion of sensor data and human feedback for personalized energy-saving recommendations," *Applied Energy,* vol. 305, 2022.

[10] K. Tsukuda and G. Masataka, "Explainable Recommendation for Repeat Consumption," *Fourteenth ACM Conference on Recommender Systems,* pp. 462-467, 2020.

[11] . A. Vultureanu-Albişi and C. Bădică, "Recommender Systems: An Explainable AI Perspective," *International Conference on INnovations in Intelligent SysTems and Applications (INISTA),* pp. 1-6, 2021.

[12] R. Shimizu, M. Matsutani and M. Goto , "An explainable recommendation framework based on an improved knowledge graph attention network with massive volumes of side information," *Knowledge-Based Systems,* vol. 239, p. 107970, 2022.

[13] S. Park, D. Chae, H. Bae, S. Park and S. Kim, "Reinforcement Learning over Sentiment-Augmented Knowledge Graphs towards Accurate and Explainable Recommendation," *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining,* pp. 784-793, 2022.

[14] B. Abdollahi and O. Nasraoui, "Transparency in Fair Machine Learning: the Case of Explainable Recommender Systems," *InHuman and machine learning,* pp. 21-35, 2018.

[15] M. Caro-Martínez, G. Jiménez-Díaz and J. Recio-García , "Conceptual Modeling of Explainable Recommender Systems:An Ontological Formalization to Guide Their Design and Development," *Journal of Artificial Intelligence Research,* vol. 71, pp. 557-589, 2021.